A New Approach for Hybrid Bayesian Networks Using Full Densities

Oliver C. Schrempf
Institute of Computer Design
and Fault Tolerance
Universität Karlsruhe
e-mail: schrempf@ira.uka.de

Uwe D. Hanebeck
Institute of Computer Design
and Fault Tolerance
Universität Karlsruhe
e-mail: Uwe.Hanebeck@ieee.org

Abstract[1]

In this article, a new mechanism is described for modeling and evaluating hybrid Bayesian networks. The approach uses Gaussian mixtures and Dirac mixtures as messages to calculate marginal densities.
The mechanism is proven to be exact, hence the accuracy of resulting marginals is only dependending on the accuracy of the conditional densities. As these densities are approximated by means of Gaussian mixtures, any desired precision can be achieved. The presented approach removes the restrictions concerning the ancestry of discrete nodes often made in literature. Hence it enables the designer to model arbitrary parent-child relationships using continuous and discrete variables.

1. Introduction

The application of Bayesian networks is evolving since it's origin in 1988 [8]. Their stochastic foundation provides a method to build models for systems with an uncertain behaviour. The common approach to model such systems is to identify parts of the system that can be represented by random variables. The behaviour of the system is then expressed by a joint probability over these random variables. The term *random variable* is usually used for scalar values only. For the sake of simplicity this paper only deals with scalar values, too, but it is easy to extend the presented approach to vector values. Hence, the term *random variable* can easily be translated into *random vector*.

Bayesian networks are considered to be an efficient representation of joint probabilities exploiting the causal background of a domain. This is achieved by representing the causal structure of a domain by a directed acyclic graph (DAG). Each random variable is depicted by a node in this graph and every edge stands for a direct dependency between two variables. Probabilistically, this dependency is expressed by a likelihood function. Bayesian networks have the big advantage, that not all possible combinations of variables and their states have to be addressed. It is sufficient to consider the conditional density of a variable given it's parents.

The first Bayesian networks were limited to a discrete domain and their likelihood functions were modeled by conditional tables. Pearl's approach to evaluate the network by means of message passing [8] was extended to continuous networks in [1]. They used Gaussian mixtures, which are sums of weighted Gaussian densities, to approximate the likelihood functions and to represent their messages.

The treatment of hybrid Bayesian networks today is mainly influenced by the articles [2–4, 7], which use so called cg-potentials. The drawback of this approach is the mere usage of the first two moments (mean and variance) to characterize continuous densities. Another problem is the method can not handle discrete nodes as children of continuous parents. An attempt to remove this restriction by using sigmod-functions is given in [6]. This approach is picked up in [5] to include it into Lauritzen's mechanism. Again, their accuracy is restricted to the first two moments of the densities.

A rarely considered problem in the context of Bayesian networks is the treatment of nonlinear dependencies between variables. The possibility to approximate the likelihood functions induced by nonlinear dependencies using Gaussian mixtures is offered in [1].

The remainder of this paper is structured as follows. The next section gives a formulation of the considered problem. Sections 3 and 4 present new formulations for hybrid conditional density functions and accordingly adapted message representations. The method for computing marginal densities given some evidence is shown in section 5 followed by the mechanism that calculates the resulting new messages in section 6. Section 7 presents an example of a hybrid Bayesian network and it's evaluation using the new approach. The corresponding results are compared to the result using Lauritzen's method.

2. Problem Formulation

Evaluating hybrid Bayesian networks requires the simultaneous consideration of continuous and discrete random variables. Hence, a compatible representation for densities in both cases has to be found. The goal of this work is to develop a mechanism that allows the evaluation of a hybrid

Proceedings of the 6th International Workshop on Computer Science and Information Technologies CSIT' 2004
Budapest, Hungary, 2004

Bayesian network in a computationally tractable way, producing a result close to the exact solution.

The drawback of using only the first two moments to describe a continuous density lies in the fact that there exist many densities having identical first moments. This can be seen in figure 1 for the functions $f_1(x) = N(x, 0, 1)$ and $f_2(x) = 0.5N(x, -\sqrt{0.5}, \sqrt{0.5}) + 0.5N(x, \sqrt{0.5}, \sqrt{0.5})$ where $N(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}$ is a Gaussian density with mean μ and deviation σ. Both densities (f_1, f_2) have mean 0 and variance 1 which are the first two moments.

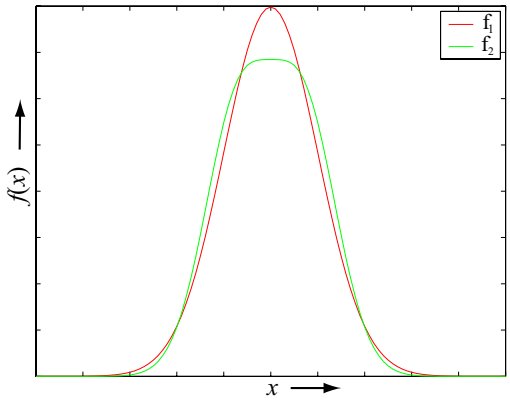

Figure 1. Two distinct densities $f_1(x) = N(x, 0, 1)$ and $f_2(x) = 0.5N(x, -\sqrt{0.5}, \sqrt{0.5}) + 0.5N(x, \sqrt{0.5}, \sqrt{0.5})$ with identical means and variances.

The simultaneous treatment of continuous and discrete variables used in our approach considers two distinct cases, which are shown in figure 2. The nodes in box shape are discrete whereas the continuous nodes have a round outline. For the parent nodes $\mathbf{u}_1, \ldots, \mathbf{u}_m$ and the child nodes $\mathbf{y}_1, \ldots, \mathbf{y}_m$ we assume a partition into continuous $(\mathbf{u}_1, \ldots, \mathbf{u}_n$ or $\mathbf{y}_1, \ldots, \mathbf{y}_n)$ and discrete $(\mathbf{u}_{n+1}, \ldots, \mathbf{u}_m$ or $\mathbf{y}_{n+1}, \ldots, \mathbf{y}_m)$ variables.

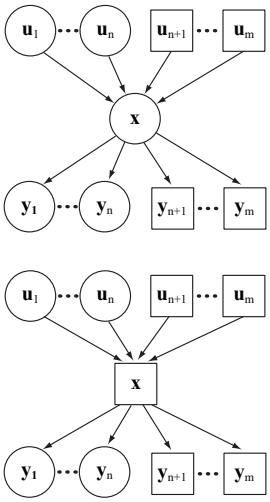

Figure 2. The simultaneous treatment of continuous and discrete variables requires the consideration of two distinct cases. The nodes in box shape are discrete, whereas the continuous nodes have a round outline.

Creating hybrid Bayesian Networks requires hybrid conditional densities to capture the relationship between continuous and discrete variables. These densities must describe the probability of a continuous or discrete random variable, depending on the state of a set of mixed parent variables. Mixed means the set of parent variables contains continuous and discrete variables as well.

Since this new approach is based on message passing, the message schemes known from pure discrete [8] or continuous [1] approaches must be extended for the use in hybrid networks. This is due to the fact that messages from continuous variables travel directly to discrete successors and vice versa. Hence, a new representation is needed, that allows simultaneous treatment of continuous and discrete densities.

3. Hybrid Conditional Densities

A hybrid conditional density $f(x|u_1, \ldots u_m)$ is given by

$$f(x|u_1 \ldots u_m) = \sum_{k_{n+1}=1}^{|\mathbf{u}_{n+1}|} \cdots \sum_{k_m=1}^{|\mathbf{u}_m|} \left(\prod_{i=n+1}^{m} \delta(u_i - k_i) \right) f^*(x|u_1, ..., u_n).$$

This formulation contains a single continuous conditional density $f^*(x|u_1, ..., u_n)$ for each joint discrete state $(u_{n+1}, \ldots, u_m)$ of $\mathbf{x}$'s discrete predecessors. The asterisk is a shorthand notation indicating the dependence on $(k_{n+1}, \ldots, k_m)$. The number of states of a discrete variable is indicated by $|\mathbf{u}_i|$. $\delta(u_i - k_i)$ is the Dirac function which can be seen as a Gaussian density with a standard deviation approaching zero

$$\delta(x - \mu) = \lim_{\sigma \to 0} N(x, \mu, \sigma) \ .$$

It has the property that it's value is zero for all $x \neq \mu$.

The conditional densities $f^*(x|u_1, ..., u_n)$ used in this paper are modeled using Gaussian mixtures in the continuous case and as sum over Gaussians and Dirac pulses in the case that $\mathbf{x}$ is discrete. This means we have a single Gaussian for each continuous parent variable and another Gaussian or sum of weighted Dirac pulses depending if $\mathbf{x}$ is continuous or discrete. This is

$$f_c^*(x|u_1, \ldots, u_n) = \sum_{j=1}^{M^*} \alpha_j^* N(x, \mu_{x,j}^*, \sigma_j^*) \cdot N(u_1, \mu_{u_1,j}^*, \sigma_{u_1,j}^*) \cdot \ldots \cdot N(u_n, \mu_{u_n,j}^*, \sigma_{u_n,j}^*)$$

in the continuous and

$$f_d^*(x|u_1, \ldots, u_n) = \sum_{j=1}^{M^*} \alpha_j^* \left(\sum_{l_j=1}^{|\mathbf{x}|} p_{l_j}^* \delta(x - l_j) \right) \cdot N(u_1, \mu_{u_1,j}^*, \sigma_{u_1,j}^*) \cdot \ldots \cdot N(u_n, \mu_{u_n,j}^*, \sigma_{u_n,j}^*)$$

in the discrete case. In the continuous case the product of Gaussians can be interpreted as a multivariate Gaussian

density with $n+1$ dimensions that is aligned with the axes of the coordinate system.

4. Messages in a Hybrid Network

The probability density over a node $\mathbf{x}$ is updated according to a set of evidence $\mathbf{e}$. As shown in figure 3, this evidence-set is divided into a subset $\mathbf{e}_{\mathbf{x}}^{+}$ of infomation from the upper part of the network and a subset $\mathbf{e}_{\mathbf{x}}^{-}$ of information from the lower part of the network according to $\mathbf{x}$.

The evidence travels to $\mathbf{x}$ by means of messages $\pi_{\mathbf{x}}(u_i)$ from the parent nodes and $\lambda_{\mathbf{y}_j}(x)$ from the child nodes. Continuous parents send their messages as Gaussian mixture densities

$$\pi_{\mathbf{x}c}(u_i) = f(u_i|\mathbf{e}_i^{+}) = \sum_{l_i=1}^{M_i} w_{l_i}^{(i)} N(u_i; \mu_{l_i,\pi}^{(i)}, \sigma_{l_i,\pi}^{(i)})$$

whereas discrete parents send a sum of weighted Dirac pulses

$$\pi_{\mathbf{x}d}(u_i) = f(u_i|\mathbf{e}_i^{+}) = \sum_{l_i=1}^{|\mathbf{u}_i|} p_{l_i}^{(i)} \delta(u_i - l_i) \ .$$

The message from a continuous child is again a Gaussian mixture

$$\lambda_{\mathbf{y}_i c}(x) = \begin{cases} \sum\limits_{l_i=1}^{M_i} w_{l_i}^{(i)} N(x; \mu_{l_i,\lambda}^{(i)}, \sigma_{l_i,\lambda}^{(i)}) & \text{if} \quad \mathbf{e}_i^{-} \neq \emptyset \\ 1 & \text{if} \quad \mathbf{e}_i^{-} = \emptyset \end{cases}$$

where $\mathbf{e}_i^{-}$ is the evidence coming from node $\mathbf{y}_i$. The message from a discrete child is a sum of weighted Dirac pulses

$$\lambda_{\mathbf{y}_i d}(x) = \begin{cases} \sum\limits_{l_i=1}^{|\mathbf{x}|} p_{l_i}^{(i)} \delta(x - l_i) & \text{if} \quad \mathbf{e}_i^{-} \neq \emptyset \\ 1 & \text{if} \quad \mathbf{e}_i^{-} = \emptyset \end{cases} .$$

The term $\mathbf{e}_i^{-} = \emptyset$ indicates, that no information was gathered in node $\mathbf{y}_i$. Hence, the message of this node is set to 1, causing no update for $\mathbf{x}$.

5. Density Update

The density[2] over $\mathbf{x}$ depending on the gathered evidence is calculated as:

$$\begin{aligned} f(x) &= f(x|\mathbf{e}) \\ &= f(x|\mathbf{e}_{\mathbf{x}}^{+}, \mathbf{e}_{\mathbf{x}}^{-}) \\ &= \alpha f(x|\mathbf{e}_{\mathbf{x}}^{+}) f(\mathbf{e}_{\mathbf{x}}^{-}|x) \\ &= \alpha \pi(x) \lambda(x) \end{aligned} \quad (1)$$

where α is a normalizing constant. The updated density over $\mathbf{x}$ is a function depending on the evidence $\mathbf{e}$ for the whole network. This evidence can be split into the evidence $\mathbf{e}_{\mathbf{x}}^{+}$ from above $\mathbf{x}$ and $\mathbf{e}_{\mathbf{x}}^{-}$ from below. Since these parts of evidence are independent, the density function over $\mathbf{x}$ can be written as a product. The density function depending on the information from above is abbreviated as $\pi(x)$ and the density depending on the information from below as $\lambda(x)$.

The information from the upper part of the net can be written as

$$\begin{aligned} \pi(x) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x|u_1, ..., u_m) \cdot \\ &\quad \prod_{j=1}^{m} f(u_j|\mathbf{e}_{\mathbf{j}}^{+}) \, \mathrm{d}u_1 \cdots \mathrm{d}u_m \ . \end{aligned}$$

This calculates the marginal density over $\mathbf{x}$ out of the information coming from every predecessor weighted by the likelihood of $\mathbf{x}$.

Inserting the definitions from above and simplifying the formula yields

$$\begin{aligned} \pi(x) &= \sum_{k_{n+1}=1}^{|\mathbf{u}_{n+1}|} \cdots \sum_{k_m=1}^{|\mathbf{u}_m|} \left(\prod_{i=n+1}^{m} \Pr(\mathbf{u}_i = k_i) \right) \\ &\quad \cdot \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f^*(x|u_1, ..., u_n) \cdot \\ &\quad \prod_{j=1}^{n} \pi_{\mathbf{x}c}(u_j) \, \mathrm{d}u_1 \cdots \mathrm{d}u_n \ . \end{aligned} \quad (2)$$

$\Pr(\mathbf{u}_i = k_i)$ is the probability, that variable $\mathbf{u}_i$ is in state k_i. (2) is equal for both continuous and discrete variables $\mathbf{x}$. To make the distinction between continuous or discrete $\mathbf{x}$ $f^*(x|u_1, ..., u_n)$ has to be chosen accordingly.

Hence, for a continuous $\mathbf{x}$ we receive the message

$$\pi_c(x) = \sum_{k_{n+1}=1}^{|\mathbf{u}_{n+1}|} \cdots \sum_{k_m=1}^{|\mathbf{u}_m|} \sum_{t=1}^{M^*} \gamma_t^* N(x; \mu_{x,t}^*, \sigma_t^*)$$

which is a Gaussian mixture density with the weights

$$\begin{aligned} \gamma_t^* &= \alpha_t^* \left(\prod_{i=n+1}^{m} \Pr(\mathbf{u}_i = k_i) \right) \cdot \\ &\quad \prod_{j=1}^{n} \sum_{l_j=1}^{M_j} w_{l_j}^{(j)} N_{u_j}(\mu_t^*; \mu_{l_j,\pi}^{(j)}, \sigma_t^* + \sigma_{l_j,\pi}^{(j)}) \ . \end{aligned} \quad (3)$$

The term $N_{u_j}(\mu_t^*; \mu_{l_j,\pi}^{(j)}, \sigma_t^* + \sigma_{l_j,\pi}^{(j)})$ describes a Gaussian Density over u_j with mean $\mu_{l_j,\pi}^{(j)}$ and variance $(\sigma_t^* + \sigma_{l_j,\pi}^{(j)})$ evaluated at μ_t^*.

In the case that $\mathbf{x}$ is discrete, the message from the upper part of the net is

$$\pi_d(x) = \sum_{k_{n+1}=1}^{|\mathbf{u}_{n+1}|} \cdots \sum_{k_m=1}^{|\mathbf{u}_m|} \sum_{t=1}^{M^*} \gamma_t^* \left(\sum_{h_t=1}^{|\mathbf{x}|} p_{h_t}^* \delta(x - h_t) \right)$$

with the same weights γ_t^* as in (3). This message is a sum of weighted Dirac pulses.

[2] Often quoted as the belief function.

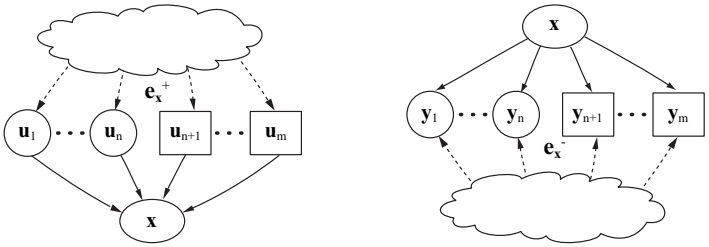

Figure 3. The evidence coming from the upper and lower part of the net is indicated by $\mathbf{e}^+$ and $\mathbf{e}^-$.

The message from the lower part of the net is written as

$$\lambda(x) \quad = \quad \prod_{j=1}^{m} \lambda_{\mathbf{y}_j}(x),$$

which is a product of the single messages coming from every child node of $\mathbf{x}$. In the case that $\mathbf{x}$ is continuous, this is again a mixture of Gaussians according to

$$\lambda_c(x) = \sum_{l_0=1}^{M_0} w'_{l_0} N(x; \mu_{l_0,\lambda}, \sigma_{l_0,\lambda})$$

with $w'_{l_0} = \prod_{j=1}^{m} w_{l_j}^{(j)}$. In the discrete case we have a product over sums of weighted Dirac pulses

$$\lambda_d(x) = \prod_{j=1}^{m} \sum_{l_j=1}^{|\mathbf{x}|} p_{l_j}^{(j)} \delta(x - l_j) \ .$$

The density function for a continuous or discrete $\mathbf{x}$ can now be obtained by multiplying the appropriate π- and λ-message.

6. Calculating the Messages to be Sent by an Updated Node

Every node receiving messages from it's neighbor sends out messages to the other neighbors as well. It sends π-messages to it's children and λ-messages to it's parents.

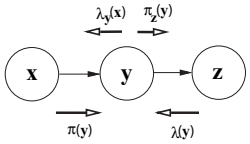

Figure 4. Messages flow into $\mathbf{y}$ to update it's density. Then $\mathbf{y}$ sends back messages to the network.

The π-Messages

The message $\pi_{\mathbf{y}_i}(x)$ that a node $\mathbf{x}$ sends to it's i-th successor is calculated the following way

$$\begin{aligned} \pi_{\mathbf{y}_i}(x) &= f(x|\mathbf{e} - \mathbf{e}_i^-) \\ &= f(x|\mathbf{e}_i^- = \emptyset) \\ &= \alpha\pi(x)\lambda(x)|_{\lambda_{\mathbf{y}_i(x)=1}}. \end{aligned}$$

This means that all evidence excluding $\mathbf{e}_i^-$, which is the information coming from $\mathbf{y}_i$, is passed ahead. Hence, this message can be calculated as shown in section 5 under the assumption $\lambda_{\mathbf{y}_i}(x) = 1$.

The λ-Messages

The calculation of the λ-messages is a little more tricky since these messages travel against the direction of the modeled dependency $f(x|u_i)$.

Depending on the continuous or discrete identity of the parent variable, the message sent by $\mathbf{x}$ is a Gaussian mixture density or a sum of weighted Dirac pulses as shown in table 1. The main information of these messages is carried

$\mathbf{u}_i$	$\lambda_{\mathbf{x}}(u_i)$
cont.	$\sum_{l_{n+1}=1}^{\lvert\mathbf{u}_{n+1}\rvert} \cdots \sum_{l_m=1}^{\lvert\mathbf{u}_m\rvert} \sum_{j=1}^{M^*} \psi_{j,i}^* N(u_i, \mu_{u_i,j}^*, \sigma_j^*)$
disc.	$\sum_{l_i=1}^{\lvert\mathbf{u}_i\rvert} \delta(u_i - l_i) \cdot \eta_{l_i}^i$

Table 1. The λ-messages from $\mathbf{x}$ to it's parent $\mathbf{u}_i$ differs for continuous or discrete $\mathbf{u}_i$.

by their weight vectors $\psi_{j,i}^*$ and $\eta_{l_i}^i$ which are calculated in different ways if $\mathbf{x}$ is continuous or discrete.

Boundary Conditions

If $\mathbf{x}$ is a root node for which no evidence is available, it's π-message is set to be the prior density for that node. This is a Gaussian mixture density for a continuous $\mathbf{x}$ and a sum of weighted Dirac pulses for a discrete $\mathbf{x}$.

If $\mathbf{x}$ is a leaf node that has not been observed so far, it's λ-message is set to 1. Hence the density for this node is calculated as $f(x) = \pi(x)$.

Exact evidence $\mathbf{x} = x_0$ is represented by $\lambda(x) = \delta(x - x_0) = N(x, x_0, 0)$. This implies $f(x) = x_0$. Uncertain evidence is expressed by means of a density.

7. Example

To evaluate the presented approach we give an example network for which some evidence is entered. Depending on that evidence the density over a node is calculated using the

approach of Lauritzen [2] and the proposed new approach. These results are compared with the true density.

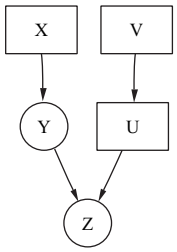

Figure 5. An example of a hybrid Bayesian network for the evaluation of the new approach. The nodes in box shape are discrete, the continuous nodes have a round outline.

Linear

The example network can be seen in figure 5. The parameters for the network in this example are the following: The two root nodes $\mathbf{x}$ and $\mathbf{v}$ have the a priori distributions $\Pr(\mathbf{x}=1)=0.4$ and $\Pr(\mathbf{x}=2)=0.6$, $\Pr(\mathbf{v}=1)=0.3$ and $\Pr(\mathbf{v}=2)=0.7$ respectively.

The discrete variable $\mathbf{u}$ has two states, hence the likelihood of $\mathbf{u}$ given $\mathbf{v}$ is modeled in the following table

$\mathbf{v}$	$\Pr(\mathbf{u}=1\|\mathbf{v})$	$\Pr(\mathbf{u}=2\|\mathbf{v})$
1	0.5	0.5
2	0.3	0.7

$\mathbf{y}$ is continuous and has a single continuous density for every state of it's discrete predecessor $\mathbf{x}$. For the sake of simplicity we choose a Gaussian density over $\mathbf{y}$ with the following parameters for every state of $\mathbf{x}$:

$\mathbf{x}$	$\mu_{\mathbf{y}}$	$\sigma_{\mathbf{y}}$
1	-2	1
2	2	1

The relationship between $\mathbf{u}$ and $\mathbf{z}$ is linear, but different for every state of $\mathbf{u}$:

$\mathbf{u}=1$	$z=2+y+w_{\mathbf{z}}$
$\mathbf{u}=2$	$z=-2+y+w_{\mathbf{z}}$

w_Z is a zero mean, Gaussian noise term, which yields the following likelihood functions:

$f(z\|y,\mathbf{u}=1)=N(z,2+y,1)$
$f(z\|y,\mathbf{u}=2)=N(z,-2+y,1)$

For using the new approach this density is approximated by the Gaussian mixture densities

$$f(z|y,\mathbf{u}=1)\approx\sum_{j=-25}^{25}\frac{1}{50}\cdot N(y,2+0.5j,1)N(z,0.5j,1)$$

and

$$f(z|y,\mathbf{u}=2)\approx\sum_{j=-25}^{25}\frac{1}{50}\cdot N(y,-2+0.5j,1)N(z,0.5j,1)\ .$$

To compare the results of the proposed approach with the exact density and the approach by Lauritzen, the evidence $\{\mathbf{v}=1,\mathbf{z}=3\}$ is entered into the network and the density over $\mathbf{y}$ is calculated. The result of this procedure is shown in figure 6.

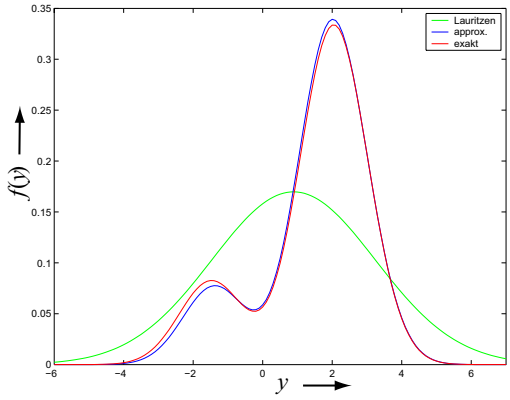

Figure 6. Given the evidence $\{\mathbf{v}=1,\mathbf{z}=3\}$ we compare the density over $\mathbf{y}$ gained by Lauritzen's method and the new approach using Gaussian mixture approximation.

Nonlinear

To show the ability of the proposed approach to deal with nonlinear relationships, the likelihood function $f(z|y,u)$ is modified to cover the following relationship

$\mathbf{u}=1$	$z=0.01\cdot y^3+w_{\mathbf{z}}$
$\mathbf{u}=2$	$z=0.01\cdot(-y)^3+w_{\mathbf{z}}$

.

The according likelihood is again approximated by Gaussian mixture densities which can be seen in figure 7.

The resulting density over $\mathbf{y}$ depending on the evidence $\{\mathbf{v}=1,\mathbf{z}=3\}$ is shown in figure 8.

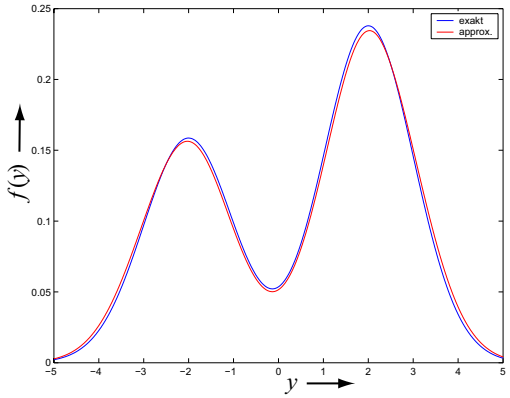

Figure 8. Comparing the exact density over $\mathbf{y}$ given the evidence $\{\mathbf{v}=1,\mathbf{z}=3\}$ with the new approach having a nonlinear dependency $f(z|x,u)$.

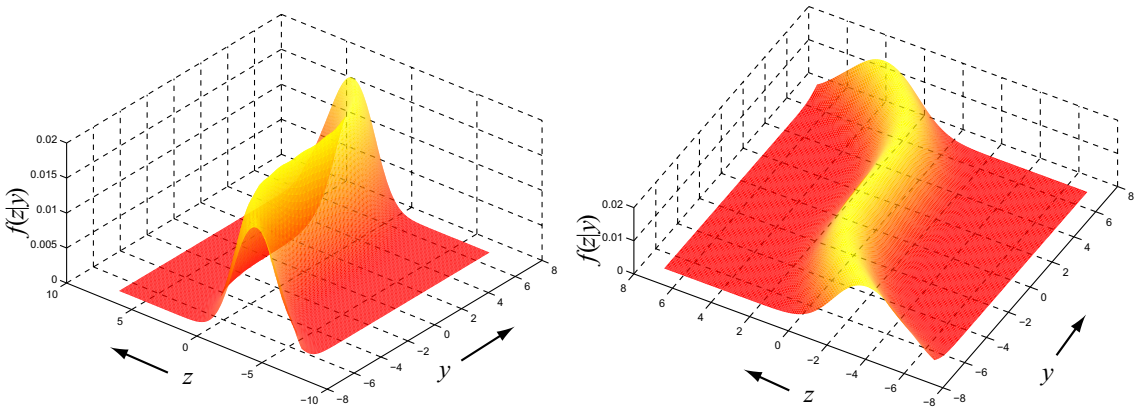

Figure 7. Nonlinear conditional density $f(z|x, \mathbf{u} = 1)$ approximated by means of a Gaussian mixture density. (two views of the same density function)

8. Conclusions

The proposed new approach for evaluating hybrid Bayesian networks allows arbitrary combinations of continuous and discrete variables while making no restrictions on their ancestry. This is achieved by the unified notation of the corresponding continuous and discrete likelihood functions. Even nonlinear dependencies between variables can be modeled by this approach.

The distance to an exact solution is only governed by the quality of the approximated likelihood functions. Hence it is possible to make a tradeoff between accuracy and complexity in computation and storage by adjusting the number of Gaussian densities used to approximate the likelihood functions.

The big improvement at this approach concerning the accuracy of the result is gained by using full densities as messages instead of only their first two moments. The usage of full densities is essential since it is not possible to reconstruct a density solely from it's first two moments.

One drawback of the approach is the limitation to singly connected graphs. One possible way to overcome this is to find a method to do clustering in the graph. Lauritzen uses clustering techniques when building his junction trees, but this results in collapsing the densities to a first two moments representation. To preserve the accuracy of our approach the persistence of the full densities in a cluster must be guaranteed.

The presented approach has already been used in the context of intention recognition in the robotics domain and showed good results. Especially the freedom in modeling and the accuracy in the evaluation were a big advantage. Since this approach is very generic, a wide field of applications is imaginable.

An interesting topic for future research is the development of effective methods for approximating likelihood functions by Gaussian mixtures. These approximations can be based upon known (non-)linear dependencies between variables or upon measured samples from an unknown underlying distribution.

References

1. Driver E., Morrell D. "Implementation of continuous Bayesian networks using sums of weighted Gaussians". In: Besnard and Hanks, (eds), *Proc. 11th Conf. Uncertainty in Artificial Intelligence*, 1995.

2. Lauritzen S. L. "Propagation of Probabilities, means and variances in mixed graphical association models". *Journal of the American Statistical Association*, 87:1098-1108.

3. Lauritzen S. L., Jensen F. "Stable Local Computation with Conditional Gaussian Distributions". Technical Report R-99-2014, Department of Mathematical Sciences, Aalborg University DK.

4. Lauritzen S. L., Wermuth N. "Graphical Models for Associations between Variables, some of which are Qualitative and some Quantitative". *The Annals of Statistics* 1989, Mar.;17(1):31-57.

5. Lerner U., Segal E., Koller D. "Exact Inference in Networks with discrete children of continuous parents". In: *Proc. of the 17th Conference on Uncertainty in Artificial Intelligence* 2001, pp. 319-238.

6. Murphy K. P. "A variational approximation for Bayesian networks with discrete and continuous latent variables". In: Blackmond-Laskey K. and Prade H. (eds) *Proc. of the Fifteenth Annual Conference on Uncertainty in Artificial Intelligence* 1999, pp. 457-466. AUAI, Morgan Kaufmann Publishers.

7. Olesen K G. "Causal Probabilistic Networks with Both Discrete and Continuous Variables". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1993, Mar.;15(3):275-279

8. Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.